DOCUMENT RESUME

ED 380 509 TM 022 878

AUTHOR Facione, Noreen C.; Facione, Peter A.
TITLE The "California Critical Thinking Skills Test" and the National League for Nursing Accreditation Requirement in Critical Thinking.
PUB DATE 94
NOTE 14p.
AVAILABLE FROM California Academic Press, 217 La Cruz Avenue, Millbrae, CA 94030.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accrediting Agencies; *Certification; College Students; Criteria; *Critical Thinking; Ethnic Groups; Higher Education; Measurement Techniques; Nurses; *Nursing Education; *Outcomes of Education; Pilot Projects; Sex Differences; *Test Construction; Test Validity
IDENTIFIERS *California Critical Thinking Skills Test (College); National League for Nursing

ABSTRACT

Criterion 20 of the National League for Nursing's accreditation process requires documentation of critical thinking (CT) as an outcome of nursing education. This raises two questions: What is meant by CT? and How can it be measured? Building on a consensus construct of CT as articulated in the American Philosophical Association 1990 Delphi Report, this paper traces the development, validation, and pilot testing of the California Critical Thinking Skills Test (CCTST). Item analysis, validity, and reliability of the CCTST are addressed, as are questions of gender, ethnicity, and native language. The pilot instrument, constructed from a pool of items developed over a 20-year research program, was administered to 1,196 college students. Both the consensus concept of CT and the CCTST instrument have applications in response to accreditation standards. (Contains 27 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

The California Critical Thinking Skills Test and the National League for Nursing Accreditation Requirement in Critical Thinking

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A Resource Paper

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
NOREEN C. FACIONE
PETER A. FACIONE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Noreen C. Facione, RN, FNP, PhD
Physiological Nursing, UC San Francisco

Peter A. Facione, PhD
Dean of Arts and Sciences
Santa Clara University

ED 380 509

California Academic Press

(c) 1994, The California Academic Press, 217 La Cruz Ave., Millbrae, CA 94030
The California Academic Press hereby grants permission for the duplication of this report for non-commercial educational uses.

TM022878

The California Critical Thinking Skills Test and the National League for Nursing Accreditation Requirement in Critical Thinking

Noreen C. Facione, RN, MSN, PhD
Department of Physiological Nursing
University of California San Francisco

Peter A. Facione, PhD,
Dean of Arts and Sciences
Santa Clara University

ABSTRACT

NLN Criterion 20 requires documentation of critical thinking as an outcome of nursing education. This raises two questions: What should we mean by "CT"? And, how can CT be measured? Building on a consensus construct of critical thinking articulated by in the American Philosophical Association 1990 Delphi Report, this paper traces the development, validation, and pilot testing of The California Critical Thinking Skills Test (CCTST). Item analysis, validity, and reliability of the CCTST are addressed, as are questions of gender, ethnicity, and native language. Both the consensus concept of CT and the CCTST instrument have applications in response to accreditation standards.

It has become a cliche that in our global economic society of galloping technological, scientific, and geopolitical change, students must learn how to learn, and learn how to think. Fact-loading memorizers who cannot analyze information, draw out the implications, evaluate the cogency of arguments, and explain how they arrived at their results will not survive in the competitive economic and political arenas of this or the next century.

Since John Dewey (1982; 1910) spoke of reflective thinking, leading educators in the United States have advocated the fostering of cognitive skills and the habits of inquiry associated with critical thinking (CT) at all levels of education. More recently Chet Meyers (1986) expressed this eloquently in Teaching Students to Think Critically when he said, "One of the aims of college education is to move students from a self-centered universe, based on limited personal experiences and concrete realities, to a richer, more abstract, realm where a multiplicity of values, visions and verities exists."

Rather than students being taught to gather soon to be obsolete facts, they are to "learn how to learn" by becoming critical thinkers. This reform agenda has been incorporated into the Department of Education's "National Education Goals for the Year 2000" Goal #5 on literacy and adult learning. Central to the achievement of this goal is the explicit objective to "assess the ability of college graduates' to think critically, to communicate effectively, and to solve problems.." (NEGR, 1991).

The National League for Nursing, through its program accreditation process, has wisely affirmed that true professionalism requires thoughtful decision making founded on the ability to make purposeful, reflective judgments which involve analysis, interpretation, inference, evaluation and explanation -- in short, to engage in critical thinking. Nursing students are to be taught these cognitive skills and nursing programs are to show evidence that their students have developed these skills. How?

Prior to the introduction of the California Critical Thinking Skills Test (Facione 1990a), there were three instruments available commercially for the assessment of CT skills at the college level: The Watson-Glaser Critical Thinking Appraisal (first developed in the 1940's, and revised most recently in 1980), the Cornell Critical Thinking Test (1985), and the Ennis-Weir Critical thinking Essay Test (1985). Stephen Norris and Robert Ennis (1989) offer sound, thorough, and readable analyses of the three CT skills test which had been published in the 1980's. J. Carter-Wells (1992) analysis, published later, includes these instruments as well as the CCTST. Carter-Wells points out that each of these instruments are based on slightly different theoretical constructs. This difference in the scope and definition of the CT construct grounding each of the four main college level CT skills

instruments introduces limits the potential for establishing concurrent validity between them.

One other instrument available for CT assessment, The California Critical Thinking Dispositions Inventory (CCTDI) (Facione & Facione, 1992) differs from the above instruments in that it does not target the measurement of CT skills. The CCTDI targets personality attributes described in the Delphi Report as characteristics of the ideal critical thinker: inquisitiveness, open-mindedness, analyticity, systematicity, confidence, truth-seeking, and maturity. Thus, the choice of any of these instruments to gather outcomes assessment data should first rest on the conceptualization of CT upon which the instrument is based. The NLN accreditation criteria call for the "nursing unit's definition of critical thinking" to be articulated. Clarifying the definition of CT, then, is the place to start CT outcomes assessment.

Kurfiss (1988) offers one of the best summaries of the development of the construct of CT in philosophy, psychology, and education prior to 1990. The result of these diverse efforts was a myriad of individual, if overlapping, definitions proposed over the decades. In 1987, as the need for a clear consensus definition of CT became increasingly apparent, the committee on Precollege Philosophy of the American Philosophical Association (APA) initiated a systematic inquiry into the current status of the construct of CT and its assessment. Using the Delphi methodology developed by the Rand Corporation (Hostrop, 1973), a facilitator conducted an anonymous, iterative, two-year, inter-communication among 46 CT experts across the United States and Canada until a consensus definition of CT was reached. The experts were drawn from Philosophy, Psychology, Education and a variety of other physical and social science disciplines. This research, which has come to be known as The Delphi Report (Facione, 1990b), represents the first consensus definition of the domain of CT. This consensus definition of CT is the conceptual basis for the CCTST.

The Delphi Report's expert consensus definition of CT represented the first time in the history of the evolution of the construct of CT that such an accord had occurred. The resulting consensus describes CT as a kind of judgment, or more specifically, "a purposeful, self-regulatory judgment." The consensus describes two dimensions of CT, the cognitive abilities dimension and the affective or dispositional dimension. Together, these two

dimensions permit the identification of the skills and sub-skills that must be cultivated to become more proficient at CT and also they permit the description of those intellectual habits which characterize persons who are adept at CT.

The APA Delphi Report's consensus statement regarding CT and the ideal critical thinker is intended as a guide to curriculum development and CT assessment (Facione, 1990b).

> "We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment was based. CT is essential as a tool of inquiry... CT is a pervasive and self-rectifying human phenomenon. The ideal critical thinker is habitually inquisitive, well-informed, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit..."

Intended as a discipline neutral description of the ideal critical thinker, the consensus description is a richly textured construct that can serve nursing well. Substituting the words "professional nurse" in the place of "ideal critical thinker" forcefully drives home the applicability of this definition to guiding nursing CT curricula. Indeed the desire to translate this rich Delphi CT construct into an assessment instrument initially motivated the development of the CCTST.

Instrument Design: A pilot instrument was constructed from a pool of 200 items developed over in 20 year research program aimed at validly and reliably testing CT (Facione & Scherer, 1978; Facione, 1973; 1984; 1986; 1987; 1989a; 1989b; Scherer & Facione 1977). Items in the 200 item pool had been previously analyzed for their ability to discriminate well between individuals and also selected for their high item-total correlations. Items selected for inclusion in the CCTST pilot instrument were chosen for their ability to

cover the domain of the five CT cognitive skills identified by the Delphi experts to lie within the CT construct: interpretation, analysis, evaluation, explanation, and inference. (A sixth skill, self-regulation, was not targeted by the CCTST because it was judged that this meta-cognitive level skill would necessarily be operative as students reflected on their answer choices throughout the testing session, and would more appropriately be directly measured by other than multiple choice items.)

Traditionally CT items have been constructed in terms of differing complexity. For instance whereas a less complex item may require that only one statement be interpreted or one inference be drawn, a more complex item may require several inferences be drawn and include distractors choices that invite more recognized errors in CT. For such items, not only must responses be identified as good or bad reasoning, but the rationale for why the reasoning is good or bad must also be identified.

Items selected for inclusion in the pilot instrument were arranged generally in order of apparent CT complexity. Each was a multiple choice item designed to be scored dichotomously, with one correct answer and three or four distractors. For instance, in the case of items targeting 'inference' each correct answer required that one make the correct inference. Some of the distractor choices were representative of frequently made errors in inference, many of which are so frequently made they are known as classical fallacies. Other distractors were designed to attract those who exhibit what are known as dispositional failures (impatience, injecting a personal bias, responding affectively, etc). Because of such distractors' attractiveness to those predisposed to commit such fallacies or display such dispositional failures, higher complexity items on the CCTST were expected to attract more incorrect responses than correct responses. As a result of this primary concern to cover this portion of the content domain, items with p-values lower than the (.4) to (.6) range normally considered to be ideal were deemed necessary inclusions in the pilot instrument.

The pool items were written using familiar topics, situations and social issues, but otherwise to be discipline-neutral and jargon-free. This item development strategy was designed to prevent advantages or disadvantages to persons who might happen to have or not have the special knowledge of any given academic discipline. Sex-role and social class

stereotypic contexts were eliminated and equal numbers of male and female referents were used in examples to decrease gender and cultural test bias.

Table 1

Sample characteristics of the CCTST pre-test post-test sample.

Gender:	Females:	N = 490	52.8%
	Males:	N = 438	47.2%
Age:	Range: 17 to 55 yrs		Mean: 22.4 years, S.D. = 5.05
University units earned:			
	Range: 0 to 170 semester units (Mean = 71, S.D. =37)		
Self-Identified Racial/Ethnic Group:			
	Native American:		N = 1
	African American:		N = 25
	Asian American:		N = 124
	Latino/Mexican American:		N = 99
	Caucasian American:		N = 533
	Other Foreign Nationals:		N = 65
	Chose not to disclose:		N = 12
	Missing		N = 86
Native Language:			
	English		N = 761 (80.9%)
	Other than English		N = 180 (19.1%)

Sample: The pilot instrument was administered to a total sample of 1196 college students at California State University Fullerton. This sample was divided into pre-test-post-test and case -control groups to permit varying instrument analyses. The largest grouping (N=945), pre-test post-test is described in Table 1. This pre-test post-test group did not differ significantly from the group at large.

Pilot Testing Procedure: Four quasi-experimental studies were conducted simultaneously to explore the attributes of the pilot instrument, only a summary of which will be discussed here due to space constraints. A more detailed report of this study can be found elsewhere (Facione, 1990c). A pretest/posttest, case/control study design was used to gather evidence for the CCTST's validity and reliability, to assess instrumentation effects and to measure gain scores after one course in critical thinking. Cases were students

enrolled in any one of four university designated courses fulfilling the campus critical thinking requirement. Controls were students who had not fulfilled the CT course requirement but were currently enrolled in the course "Introduction to Philosophy." In all 37 class sections and 20 professors participated in the study. In all 1673 individual completed tests were available for evaluation. This number of completed instruments provided more than adequate power for all subsequent statistical analyses.

The pilot CCTST was administered under conditions similar to those in which the final instrument was intended for use, namely, in college level classrooms, within a 45 minute time frame. The CCTST was not used as either a requirement or a grade for any student, rather they were asked to complete the test voluntarily. This planned study design was felt to potentially underestimate the true gain score between pretest and posttest as students were more likely to try harder on the pretest, given in the first week of the semester than on the posttest which was given independent of their grades for the semester. Students were given no advanced notification of test administration, and were told vaguely that their cooperation was appreciated as part of a much larger University research effort regarding the campus CT requirement. Concerns regarding poor student cooperation for participating in the study proved unfounded, with 95% of those invited to participate completing CCTST instruments for analysis.

Results: The item analyses and statistical analyses were run using ParSCORE 2.0 (1988) and SPSS 2.0 (1988) respectively. CCTST total scores ranged from 2 to 29 in the pretest group, and ranged from 3 to 31 in the post-test sample. Total score distributions for all study samples approximate the normal distribution. One item was observed to have a somewhat poorer item-total correlation, to discriminate more poorly between individuals and to have no effect on the KR-20 reliability. For these reasons, and to decrease respondent burden, given that not all students completed the instrument in the allotted time, the item was dropped from the pilot. All other items were retained in the final 34 item CCTST instrument.

Validity and Reliability: The content validity of the CCTST rests on its relationship to the APA Delphi Report research. Consideration of concurrent validity in an instrument

such as the CCTST must first address the question of what external criterion we would wish to predict. Evidence for concurrent validity of the CCTST connect CCTST scores with other measures of college students' aptitude and achievement. Total scores of the pretest group (so as not to confound effects of CT instruction) correlate significantly with college level grade point average (.200, $p<.001$), SAT verbal (.550, $p<.001$) SAT math (.439, $p<.001$), and Nelson-Denny Reading scores (.491, $p<.001$), which are themselves described as predictors of freshman level college grade point average. Construct validity of the CCTST is supported by results of the pretest-post-test measure of significant gains in cases but not in controls (Facione, 1990c), as well as by the high and significant correlation ($r = .667$) being reported between the CCTST and the CCTDI being reported in several pilot and study samples (Facione, Sanchez & Facione, 1994; Facione, Facione & Sanchez, 1994).

In terms of predictive validity, clearly what one would wish to predict is the practice of critical thinking in a given setting, not at all an easy criterion to measure by any known means. Evidence for predictive validity of the CCTST awaits the completion of longitudinal cohort studies.

The Kuder-Richardson internal reliability coefficients computed for each of the sections of the divided sample ranged from .68 -.69. This internal consistency estimate of reliability deserves particular interpretation, as we are accustomed to higher levels in instruments measuring narrower, single focus domain concepts. In non-homogeneous instruments aimed at testing a broad range of a complex construct, in instruments where items are intended to discriminate well between individuals, and on instruments which rely on dichotomous scoring (Nunnally, 1978), an achievable level of internal reliability in such instruments is typically regarded to be .65 -.75 (Norris & Ennis, 1989). Under this criteria the KR-20 of .68-.69 supports its reliability to measure CT skills. One approach to increasing internal consistency reliability, of course, would be to increase test length. Using the Spearman-Brown Prophesy Formula, given this average inter-item correlation of (.06), increasing the number of test items to approximately 62 similar items might be expected to increase the reliability coefficient to approximately .80. This potential change in the CCTST would be unfeasible, however, in light of its intended use for curriculum evaluation or student placement within the typical classroom time period. Further, in contrast to the

Watson-Glaser and the Cornell instruments, the complexity of items on the CCTST creates sufficient mental fatigue that increasing the test length to 62 items would likely decrease the overall reliability estimate of true scores in terms of increasing error due to fatigue.

The APA Delphi Report indicates that the various cognitive skills involve in CT do not operate as independent or isolated factors, but rather in an interdependent and interconnected way. Therefore, a factor analysis aimed at parsing out the differences between the skills of inference, analysis, and evaluation can be predicted to fail, if the instrument being tested has succeeded in requiring that these skills be used interactively to respond to an individual item. For this same reason, although items are identified as targeting the particular skill areas of analysis, evaluation and inference, scores for these subscales are not independent and should probably not be used for more than gross indicators of possible CT strengths and weaknesses.

Significant gains in CCTST total score were observed in the case group as compared to the control group. This was true without consideration of which individual CT course was taught (Psychology, Reading or Philosophy) and true regardless of who the CT instructor was. Neither age nor number of semester units completed were found to be significant predictors of CCTST total score. These two reported findings were central to disconfirming two of the experimental hypotheses: that CT skills were improved by university coursework in general, and that critical thinking ability improved with increasing age in general.

The difference in CCTST total scores by gender was not significant at the $p < .05$ level of probability, although the overall mean scores for males (16.3) was higher than that of females (15.9). Perhaps of particular interest to nursing programs given the large numbers of females in their student cohort, gain scores were significant by gender ($p < .013$) with males showing a significantly larger gain (1.2 overall) than females (0.4 overall). In light of the fact that females in the sample had generally higher college grade point averages (mean = 2.75) than males (mean = 2.64), this raises the question for future research of whether males might be advantaged in the traditional pedagogical approaches used to teach CT in the classroom.

Overall race and ethnicity per se were not significant determinants of gain scores on the CCTST, but native English speaking ability was significantly associated with larger gain scores ($p < .002$). Whereas the African American students (admittedly a small portion of the sample) showed higher average gain scores (2.0) than the overall sample, Asian (-0.1) and Latino (0.2) showed no significant gains overall, each of these two groups including large numbers of students for whom English was a second language. Controlling for other factors, no differences in total scores were observed by academic major, supporting the claim for the instrument's content neutrality.

Alternate Forms of the Instrument: In 1992 an alternate form of the CCTST was introduced. A description of its validation study is reported elsewhere (Facione, 1992). Form B of the CCTST was developed to be, to the extent possible, both conceptually and statistically equivalent to From A (KR-20 = .69). In terms of the internal logic of each question and answer choice, and in terms of a paradigm analysis of the CT required to derive the designated answer, the two forms are parallel, question for question. In terms of the length of each item, the position of items on pages, and overall order of items on the tests, Form B parallels Form A. Form B contains 21 new items and 13 retained from Form A. For these 13, the order of the answer has been scrambled. In content, the Form B item stems range over the same kinds of familiar issues, topics and situations used in From A.

Certainly, in responding to the NLN accreditation requirements, multiple methods of assessment are preferable for a thorough CT assessment plan, and the use of a CT skills test can be one method of gathering useful data regarding CT outcomes. Attitudinal inventories, essay tests, case study analyses, theoretical debates, role playing, talk aloud exercises, analyses of decision making in clinical practice settings, etc. provide opportunities for CT assessment by trained observers who are focusing specifically on CT skills and dispositions.

The CCTST offers one method of assessment based on a clearly articulated construct of CT, the Delphi Report's consensus definition, a construct that is being endorsed by an ever growing community of scholars with interest in CT assessment. Although limited for the assessment of students whose native language is other than English, thus far evidence

of gender and ethnicity bias has not been observed for this instrument. Content neutrality is a strength of this instrument, presenting no advantage to students from any one discipline. Although group norms for the instrument are available (Facione, 1990a), since scoring and data are controlled by the users at the testing site, the development of preferable local norms is possible. Control of data also more easily permits creative assessment programs and the longitudinal study of student cohorts. Demonstrated gains in CT skills measured by the CCTST support its use for outcomes assessment plans and for accreditation or program review purposes where aggregate information about students at various program levels -- for example, at entry and at exit -- can contribute to an overall evaluation of the program's effectiveness.

Documentation for accreditation purposes can focus on input, process, or outcomes. At the input level, program goals, course objectives, and syllabi attest to the intentions to develop the level of CT requisite for successful professional practice. At the process level, the program faculty can describe their methods of instruction, pedagogues, and exercises which used to foster CT. In asking for documentation and evidence, the NLN rightly assumes that while we value CT in our students, we may not always be instilling it in them. Whether in response to the NLN or to one's concern for teaching effectiveness, only outcomes assessment, using independently validated instrumentation, provides direct evidence that those instructional aspirations and pedagogical strategies are issuing in the further refinement and development of critical thinking skills in students.

In 1994 Dr. Noreen Facione began a national critical thinking meta-study the purpose of which is to aggregate student assessment data on a variety of variables and measures, including the CCTST and CCTDI, and hence, to assist individual institutions in the interpretation of their own data. The focus of this meta-study is on programs in nursing and other heath professions. At the time of this writing, 45 institutions had joined the study. For information, contact The California Academic Press. Phone/FAX (415) 697-5628.

References

Carter-Wells, J. (1992). The Carter-Wells Matrix. Inquiry: Critical Thinking Across the Disciplines. Vol 9, (1), 13.

Dewey, J. (1982, Originally published 1910). How We Think. Lexington, MA: Heath.

Ennis, R. H., Millman, J., & Tomko, T. N. (1985). Cornell Critical Thinking Tests, Levels X and Z. Pacific Grove, CA: Midwest Publications.

Ennis, R. H., & Weir, E. (1985). The Ennis-Weir Critical Thinking Essay Test. Pacific Grove, CA: Midwest Publications.

Facione, N. C., Facione, P. A. & Sanchez, C. A. (1994). Journal of Nursing Education, 33(8), 345-350. The disposition toward CT as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory (CCTDI).

Facione, P.A. (1973). An Introduction to Logic and Logical Thinking. Dubuque, Iowa: Kendall/Hunt.

Facione, P.A. (1984). Toward a theory of critical thinking. Liberal Education, 70(3), 253-261.

Facione, P.A. (1986). Testing college level critical thinking. Liberal Education, 72(3), 221-231.

Facione, P.A. (1987). Teaching about fallacies. Teaching Philosophy, 10, 211-217.

Facione, P.A. (1989a). A partial critical thinking bibliography with emphasis on testing. ERIC Doc. No. TM 013578.

Facione, P.A. (1989b). Assessing inference skills. ERIC Doc. No. TM 012917.

Facione, P.A. (1990a; 1992). The California Critical Thinking Skills Test: (CCTST). Forms A (1990) and Form B (1992). Millbrae, CA.: California Academic Press.

Facione, P.A. (1990b). The Delphi Report. Critical Thinking: A statement of expert consensus for purposes of educational assessment and instruction. ERIC Doc. No. TM 014423.

Facione, P.A. (1990c). Technical Reports #1-4. ERIC Docs. Nos. TM 015818, TM 015819, TM 015962, TM 015996.

Facione, P.A. (1991). Critical Thinking Analyses of Eight CCTST Items, Millbrae, CA.: California Academic Press.

Facione, P.A., and Facione, N., (1992). The California Critical Thinking Dispositions Inventory (CCTDI). Millbrae, CA.: The California Academic Press.

Facione, P.A., Sanchez, C.A. & Facione, N.C. (1994). Are College Students Disposed to Think?. ERIC Doc: ED 368-311.

Facione, P.A. & Scherer, D. (1978). Logic and Logical Thinking: A Modular Approach. New York: McGraw-Hill. 2nd Ed. Woodbridge, CT: Oxbow Press.

Hostrop, R. (1973). The Delphi Technique. Managing Education For Results. Homewood, IL: ETC Publishing.

Kurfiss (Gainen), J. (1988) .Critical Thinking: Theory, Research, Practice, and Possibilities ASHE-ERIC Higher Education Report 2.

National League for Nursing. (1989) NLN Self-Study Manual: guidelines for preparation of the self-study report, Council of Baccalaureate and Higher Degree Programs, NLN.

National Educational Goals Resources Group (NERG). (1991). Interim Report. US Department of Education, Office of Educational Statistics and Research.

Meyers, Chet. (1986). Teaching Students to Think Critically. (p. 27). San Francisco, CA: Jossey-Bass Publishing.

Norris, S. P. & Ennis, R. H. (1989). Evaluating Critical Thinking. (p. 46f). Pacific Grove, CA: Midwest Publications.

Nunnally, J. C. (1978). Psychometric Theory. (p. 208-231). New York: McGraw-Hill.

Scherer, D. & Facione, P. A. (1977). A pre/post test for introductory logic. Metaphilosophy, 7(4), 342-347.

Watson, G. & Glaser, E.M. (1980). Watson-Glaser Critical Thinking Appraisal, Forms A and B. San Antonio: Psychological Corporation.